

I CLAIM:

1. A method for creating automated inferences, comprising:

(a) extracting a database record from a structured literature database;

(b) parsing the database record to extract one or more individual information
5 fields, wherein the one or more individual information fields include a set of chemical or
biological molecule names;

(c) filtering the extracted set of chemical or biological molecule names to create a
filtered set of chemical or biological molecules names;

(d) determining whether a chemical or biological molecule name from the filtered
10 set has been stored in an inference database,

and if not,

storing the chemical or biological name in the inference database,
and setting a co-occurrence count to a starting value for each pair of names including the
chemical or biological name and other names from the filtered set that the chemical or
15 biological name co-occurs with;

and if so,

incrementing co-occurrence counts for each pair of chemical or biological
names including the chemical or biological name;

(e) repeating steps (a)-(d) for unique database records in the structured literature
20 database;

(f) optionally constructing a connection network using a plurality of database
records from the inference database including co-occurrence counts;

(g) applying one or more analysis methods directly to database records in the inference database or to the optional connection network to determine possible inferences
25 of physico-chemical relationships between chemical or biological molecules; and

(h) generating automatically a plurality of inferences regarding physico-chemical relationships between chemical or biological molecules using the results from the one or more analysis methods.

2. The method of Claim 1 further comprising a computer readable medium having stored therein instructions for causing a processor to execute the steps of method.

3. The method of Claim 1 wherein extracting step includes extracting a plurality of database records with a pre-determined database record structure.

4. The method of Claim 3 wherein the extracting step includes extracting a database record with a pre-determined structure from Medline, PubMed, Biological Abstracts or Science Citation Index databases.

5. The method of Claim 1 wherein the parsing step includes parsing the database record to extract a record information field indicating two or more chemical or biological molecule names used in an experiment recorded in the database record.

6. The method of Claim 1 wherein the filtering step includes filtering the chemical or biological molecule names against a list of trivial chemical or biological molecule names to be ignored.

7. The method of Claim 1 wherein the step of optionally constructing a connection network includes constructing a connection network including a plurality of nodes representing a plurality of chemical or biological molecules names and a plurality of arcs connecting the plurality of nodes, wherein the plurality of arcs represent co-
5 occurrences between chemical or biological molecules.

8. The method of Claim 1 wherein the applying step includes applying statistical analysis methods to co-occurrence counts stored in the inference database.

9. The method of Claim 1 wherein the generating step includes generating automatically inferences for physico-chemical interactions between chemical or biological molecules using the co-occurrence counts stored in the inference database.

10. The method of Claim 9 wherein the physico-chemical interactions between chemical or biological molecules include physico-chemical interactions for chemical or biological molecules for cells.

11. The method of the Claim 1 wherein the chemical or biological molecule names include natural or synthetic chemical compound or chemical molecule names or natural or synthetic biological molecule or biological compound names.

12. The method of Claim 1 further comprising storing the plurality of inferences in the inference database.

13. The method of Claim 1 further comprising applying subsequent analysis methods to the connection network to reject trivial inference associations.

14. The method of Claim 13 wherein the subsequent analysis methods include assigning derived numerical values to arcs in the connection network based on co-occurrence counts, assigning derived numerical values to arcs in the connection network based on analysis of a temporal pattern of an inference association's co-occurrence count
5 as a function of another variable, conducting a mutual information analysis, or conducting a Citation analysis.

15. The method of Claim 1 wherein the step incrementing step includes incrementing a plurality of co-occurrence counts for pairs of chemical or biological molecule names in the filtered set.

16. A method for checking automatically created inferences, comprising
creating a connection network from an inference database including inference
knowledge, wherein the connection network includes a plurality of nodes representing a
plurality of chemical or biological molecules names and a plurality of arcs connecting the
5 plurality of nodes, wherein the plurality of arcs represent co-occurrences counts between
chemical or biological molecules and wherein the inference database includes a plurality
of inference database records including inference association information;

applying one or more analysis methods to the connection network to determine
any trivial inference associations; and

10 deleting automatically database records determined to include trivial inference
associations from the inference database, thereby improving the inference knowledge
stored in the inference database.

17. The method of Claim 16 further comprising a computer readable medium
having stored therein instructions for causing a processor to execute the steps of method.

18. The method of Claim 16 wherein the applying step includes assigning derived
numerical values to arcs in the connection network based on co-occurrence counts,
assigning derived numerical values to arcs in the connection network based on analysis of
a temporal pattern of an inference association's co-occurrence count as a function of
5 another variable, conducting a mutual information analysis, or conducting a Citation
analysis.

19. The method of Claim 16 wherein the inference association information includes physico-chemical interactions for chemical or biological molecules for cells.

20. The method of Claim 16 wherein the connection network includes a directed graph or an un-directed graph.

21. An automated inference system, comprising, in combination:

an automated inference creator for extracting a database record from a structured literature database, parsing the database record to extract one or more individual information fields, wherein the one or more individual information fields include a set of
5 chemical or biological molecule names, filtering the extracted set of chemical or biological molecule names to create a filtered set of chemical or biological molecules names, determining whether a chemical or biological molecule name from the filtered set has been stored in an inference database, and if not, storing the chemical or biological name in the inference database, and setting a co-occurrence count to a starting value for
10 each pair of names including the chemical or biological name and another name from the filtered set that the chemical or biological name co-occurs with, and if so, incrementing co-occurrence counts for each pair of chemical or biological names including the chemical or biological name, optionally constructing a connection network using a plurality of database records from the inference database including co-occurrence counts,
15 applying one or more analysis methods directly to database records in the inference database or to the optional connection network to determine possible inferences of

physico-chemical relationships between chemical or biological molecules, and generating automatically a plurality of inferences regarding physico-chemical relationships between chemical or biological molecules using the results from the one or more analysis
20 methods;

an automated inference checker for creating a connection network from an inference database including inference knowledge, wherein the connection network includes a plurality of nodes representing a plurality of chemical or biological molecules names and a plurality of arcs connecting the plurality of nodes, wherein the plurality of
25 arcs represent co-occurrences counts between chemical or biological molecules and wherein the inference database includes a plurality of inference database records including inference association information, applying one or more analysis methods to the connection network to determine any trivial inference associations, deleting automatically database records determined to include trivial inference associations from
30 the inference database, thereby improving the inference knowledge stored in the inference database;

one or more connection networks for creating inferences, wherein a connection network includes a plurality of nodes representing a plurality of chemical or biological molecules names and a plurality of arcs connecting the plurality of nodes, wherein the
35 plurality of arcs represent co-occurrences between chemical or biological molecule names in indexed scientific literature database records; and

an inference database for storing co-occurrence information, generating automatically inferences regarding known physico-chemical interactions regarding

chemical or biological molecules using the co-occurrence counts stored in the inference
40 database.

22. The system of Claim 21 wherein the physico-chemical interactions regarding chemical or biological molecules include physico-chemical interactions for chemical or biological molecules for cells.

23. The system of Claim 21 wherein the connection network includes an un-directed graph or a directed graph.

McDONNELL BOEHNEN
HULBERT & BERGHOFF
300 SOUTH WACKER DRIVE
CHICAGO, ILLINOIS 60606
TELEPHONE (312) 913-0001